

AD-A175 360

ON INACCURACIES IN OMEN'S APPROXIMATION FOR THE
BAYESIAN ABILITY ESTIMATE(U) CENTER FOR NAVAL ANALYSES
ALEXANDRIA VA D R DIVGI FEB 86 CRM 86-46

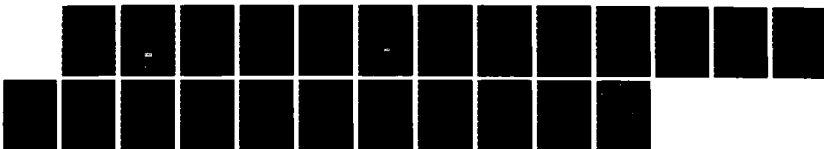
1/1

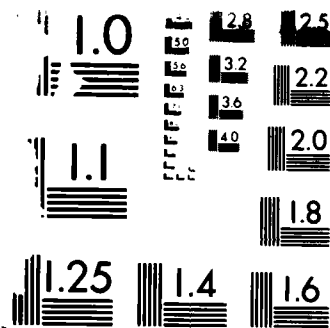
UNCLASSIFIED

N00014-83-C-0725

F/G 5/10

NL





U.S. GOVERNMENT PRINTING OFFICE: 1963 O - 348-100
 NATIONAL BUREAU OF STANDARDS

12

RESEARCH MEMORANDUM

ON INACCURACIES IN OWEN'S APPROXIMATION FOR THE BAYESIAN ABILITY ESTIMATE

D. R. Divgi

AD-A175 360

DTIC FILE COPY

A Division of

CNA

Hudson Institute

CENTER FOR NAVAL ANALYSES

4401 Ford Avenue • Post Office Box 16268 • Alexandria, Virginia 22302-0268 • (703) 824-2000

DTIC
ELECTE
DEC 29 1986
S E D

AD-A175366

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified		1b. RESTRICTIVE MARKINGS	
2a. SECURITY CLASSIFICATION AUTHORITY		3. DISTRIBUTION / AVAILABILITY OF REPORT	
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE		Approved for Public Release; distribution unlimited.	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) CRM 86-46		5. MONITORING ORGANIZATION REPORT NUMBER(S)	
6a. NAME OF PERFORMING ORGANIZATION Center for Naval Analyses	6b. OFFICE SYMBOL (If applicable) CNA	7a. NAME OF MONITORING ORGANIZATION Deputy Chief of Staff (RDS)	
6c. ADDRESS (City, State, and ZIP Code) 4401 Ford Avenue Alexandria, Virginia 22302-0268		7b. ADDRESS (City, State, and ZIP Code) Headquarters, Marine Corps Washington, DC 20380	
8a. NAME OF FUNDING / ORGANIZATION Office of Naval Research	8b. OFFICE SYMBOL (If applicable) ONR	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-83-C-0725	
8c. ADDRESS (City, State, and ZIP Code) 800 North Quincy Street Arlington, Virginia 22217		10. SOURCE OF FUNDING NUMBERS	
		PROGRAM ELEMENT NO. 65154N	PROJECT NO. R0148
		TASK NO.	WORK UNIT ACCESSION NO.
11. TITLE (Include Security Classification) On Inaccuracies in Owen's Approximation for the Bayesian Ability Estimate			
12. PERSONAL AUTHOR(S) D. R. Divgi			
13a. TYPE OF REPORT Final	13b. TIME COVERED FROM Feb 1986 TO Mar 1986	14. DATE OF REPORT (Year, Month, Day) February 1986	15. PAGE COUNT 22
16. SUPPLEMENTARY NOTATION			
17. COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)	
FIELD	GROUP	SUB-GROUP	
19. ABSTRACT (Continue on reverse if necessary and identify by block number) <p>In computerized adaptive testing, an examinee's ability is often estimated using Bayesian procedures. Its calculation is greatly simplified if one uses Owen's approximation. However, when the test length is moderate (e.g., 15 items), the accuracy of Owen's approximation is inadequate for calculating the examinee's score. This is illustrated using the nine power subtests in the Armed Services Vocational Aptitude Battery. The recommended procedure is to use Owen's approximation for adaptive item selection, and to compute the exact Bayesian estimate after the entire test has been administered.</p>			
20. DISTRIBUTION / AVAILABILITY OF ABSTRACT <input type="checkbox"/> UNCLASSIFIED/UNLIMITED <input checked="" type="checkbox"/> SAME AS RPT <input type="checkbox"/> DTIC USERS		21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL Lt. Col. R. W. Geary		22b. TELEPHONE (Include Area Code) 694-4165	22c. OFFICE SYMBOL DC/S

CENTER FOR NAVAL ANALYSES

J. Biol. Chem. • Volume 269, Number 18, September 2, 1994

18 March 1986

MEMORANDUM FOR DISTRIBUTION LIST

Subj: Center for Naval Analyses Research Memorandum 86-46

Encl: (1) CNA Research Memorandum 86-46, "On Inaccuracies in Owen's Approximation for the Bayesian Ability Estimate," February 1986

1. Enclosure (1) is forwarded as a matter of possible interest.
2. The Department of Defense plans to implement a computerized adaptive testing (CAT) version of the Armed Services Vocational Aptitude Battery (ASVAB) in the near future. An important issue for the CAT program is the means whereby an individual's score is calculated from his responses to CAT-ASVAB test questions. This Research Memorandum examines the accuracy of the method planned for use in the CAT-ASVAB system.

William A. Jones

William H. Sims
Director
Marine Corps Manpower
and Training Program

Distribution List:
Reverse Page

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Classification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

Subj: Center for Naval Analyses Research Memorandum 86-46

Distribution List

Deputy Chief of Staff, Manpower (3 copies)

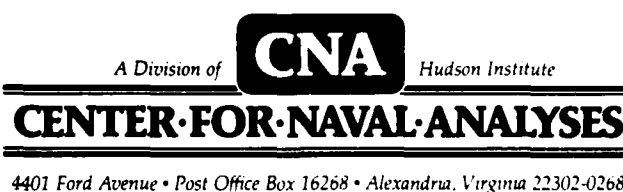
Deputy Chief of Staff, Research, Development and Studies (2 copies)

CAT-ASVAB Psychometric Committee (8 copies)

ON INACCURACIES IN OWEN'S APPROXIMATION FOR THE BAYESIAN ABILITY ESTIMATE

D. R. Divgi

Marine Corps Operations Analysis Group



ABSTRACT

In computerized adaptive testing, an examinee's ability is often estimated using Bayesian procedures. Its calculation is greatly simplified if one uses Owen's approximation. However, when the test length is moderate (e.g., 15 items), the accuracy of Owen's approximation is inadequate for calculating the examinee's score. This is illustrated using the nine power subtests in the Armed Services Vocational Aptitude Battery. The recommended procedure is to use Owen's approximation for adaptive item selection, and to compute the exact Bayesian estimate after the entire test has been administered.

TABLE OF CONTENTS

	<u>Page</u>
List of Tables	v
Introduction	1
Method	1
Data Set	2
Calculations	2
Results	3
Implications	4
References	14

LIST OF TABLES

	<u>Page</u>
1 Results for Subset GS	5
2 Results for Subset AR	6
3 Results for Subset WK	7
4 Results for Subset PC	8
5 Results for Subset AI	9
6 Results for Subset SI	10
7 Results for Subset MK	11
8 Results for Subset MC	12
9 Results for Subset EI	13

INTRODUCTION

The Department of Defense plans to try out computerized adaptive testing (CAT) of the Armed Services Vocational Aptitude Battery (ASVAB) in the near future. Adaptive testing involves the use of a large item pool. For any given examinee, the next item to be administered is selected on the basis of the current ability estimate. Bayesian ability estimation is commonly used for this purpose. When testing begins, the ability distribution in the population provides the prior information about each examinee's ability. As each item is presented and answered, the ability estimate is updated. The Bayesian estimate is the mode of the posterior distribution which, except for a multiplicative factor, is the product of the prior distribution and likelihood functions of the examinee's item responses.

The exact Bayesian estimate requires extensive computations. These add to the time the examinee must wait before the next item is presented. Since long delays are undesirable, an approximation is used to calculate the ability estimate.

The approximation to be used in the CAT-ASVAB project is due to Owen [1]. The prior ability distribution is assumed to be normal, and the likelihood functions of item responses are given by the three-parameter normal ogive model. After each item, Owen replaces the true posterior distribution by a normal distribution with the same mean and variance. This greatly simplifies the calculations but also introduces errors of unknown sign and magnitude.

This Research Memorandum presents results based on real data, using power subtests in the ASVAB. Owen's approximation was compared with the correct Bayesian modal estimate. Results show that the errors are not small enough to be ignored.

METHOD

Analyses are based on real data from the Joint Services CAT Validity Study [2].

Data Set

The data set consists of item responses from about 7,000 recruits. Each examinee was administered 15 items (10 in case of PC), selected for maximum information at the current ability estimate. Estimates calculated by Owen's procedure were available on tape, as were code numbers of the items and the examinees' responses (scored 1 if correct, 0 if wrong). The prior distribution was assumed to have zero mean and unit variance. Item parameters used in the calculations were taken from a separate tape.

The tests analyzed were General Science (GS), Arithmetic Reasoning (AR), Word Knowledge (WK), Paragraph Comprehension (PC), Auto Information (AI), Shop Information (SI), Math Knowledge (MK), Mechanical Comprehension (MC), and Electronics Information (EI).

Calculations

The exact Bayesian modal estimate was computed for each examinee. In keeping with the original calculations, standard normal prior and normal ogive model were used.

Owen's mean and variance estimates for the posterior distribution were used to guide the search for the true mode. Beginning with Owen's mean, and step size equal to a tenth of the posterior sigma, successive intervals were tested to see if they contained the true mode. When the desired interval was found, the mode was computed by successive approximations based on quadratic interpolation. Iteration was stopped when the estimate changed by less than 0.01. The error in Owen's approximation was defined as Owen's estimate minus the true mode.

As the error distribution is likely to depend on the ability level, it was quantified as Owen's estimate rounded to one decimal. Mean and standard deviation of errors were calculated separately at each ability level. As pointed out by Sympton [3] in connection with the bias in Bayesian estimates, effects of nonzero averages are likely to disappear after equating. The real problem is that the error varies from one person to another, which creates the possibility of unfairness to some examinees.

At each ability level, "adjusted" errors were calculated by subtracting out the mean value. These were used to obtain percentages of examinees for whom adjusted error was above 0.1 or below -0.1 units of ability, on a scale in which the prior distribution has unit variance.

The true Bayesian estimate depends on which items were presented, and how the examinee answered them. It does not depend on the order in which the items are introduced into the computations. Such is not the case with Owen's approximation. To illustrate this point, Owen's procedure was applied to the same data, but the order of items was reversed. (The last item was added to the prior distribution, then next to the last item, and so on.) The "order effect" was defined as this new estimate minus the original estimate. Its mean and standard deviation were computed separately at each ability level.

RESULTS

The results are presented in Table 1 through 9. Only those ability levels with at least 25 examinees are shown. Since the standard deviation of the ability distribution is 1, an error of 0.1 corresponds to roughly one standard score point on the usual ASVAB scale.

The mean and sigma of error vary with ability level. The spread of errors is more important than the mean. Its implications can be seen in the fifth column. Use of Owen's approximation inflates the standard score by one point or more for a nontrivial fraction of the examinees. Unfair advantage for some persons amounts to unfair disadvantage for all others. Some subtests also reveal significant percentages of adjusted errors below -0.1.

The "order effect" has substantial mean and spread. It implies that, if two examinees provide exactly the same answers on the same items but in different orders, they are likely to get different scores from the Owen procedure.

Percentages of large errors are higher for some subtests than for others. In the case of PC, this may be due to the test length being 10 rather than 15 items. In the case of Auto and Shop Information subtests, the reason may be that item selection was based on item parameter estimates obtained from a small sample (N=300). Although better estimates were obtained later and Owen calculations repeated, the match between examinee and items remained unsatisfactory. It is not clear why MC has a higher percentage of large errors than the remaining five.

IMPLICATIONS

The results make it clear that some examinees will benefit and others will suffer if the final ability estimate is Owen's approximation rather than the true mode of the posterior distribution. One cannot argue that the errors involved are small. Even if they were, such an argument would be justified only if the inequity was unavoidable, e.g., being due to sampling errors. That is not the case here. The errors result from using an approximation when exact calculations are feasible.

There is no reason to question the use of Owen's approximation while selecting items. A few items may be different if the true mode is used. However, use of the approximation will result in no more than a slight reduction in the reliability of the test. It will not lead to any systematic errors. It is only the final ability estimate, at the end of the test, that needs to be recomputed from the true posterior distribution.

Table 1

RESULTS FOR SUBTEST GS*

Owen's Estimate	N	Errors		Adjusted Errors		Order Effect	
		Mean	S.D.	%Above 0.1	%Below -0.1	Mean	S.D.
-1.0	25	-0.07	0.033	0.0	0.0	-0.13	0.116
-0.9	55	-0.07	0.038	3.6	0.0	-0.11	0.123
-0.8	59	-0.06	0.072	6.8	0.0	-0.13	0.132
-0.7	108	-0.06	0.065	6.5	0.0	-0.10	0.118
-0.6	126	-0.07	0.050	5.6	0.0	-0.09	0.129
-0.5	128	-0.06	0.046	0.8	0.0	-0.08	0.094
-0.4	181	-0.06	0.028	0.6	0.0	-0.10	0.106
-0.3	229	-0.05	0.041	2.2	0.0	-0.08	0.115
-0.2	261	-0.04	0.042	1.9	0.0	-0.09	0.134
-0.1	328	-0.04	0.045	3.4	0.0	-0.13	0.137
0.0	329	-0.04	0.050	4.0	0.0	-0.15	0.131
0.1	378	-0.04	0.053	5.8	0.0	-0.16	0.118
0.2	361	-0.04	0.056	4.7	0.0	-0.14	0.111
0.3	419	-0.03	0.053	3.6	0.0	-0.15	0.124
0.4	459	-0.02	0.061	5.9	0.0	-0.17	0.133
0.5	528	-0.02	0.057	5.7	0.0	-0.18	0.147
0.6	522	-0.03	0.060	6.5	0.0	-0.21	0.150
0.7	465	-0.02	0.068	8.4	0.0	-0.22	0.146
0.8	434	-0.02	0.052	5.8	0.0	-0.22	0.151
0.9	406	-0.01	0.055	6.2	0.0	-0.22	0.169
1.0	347	-0.01	0.057	6.6	0.3	-0.21	0.187
1.1	307	0.00	0.062	6.5	0.0	-0.20	0.152
1.2	243	0.00	0.059	6.2	0.0	-0.21	0.224
1.3	227	-0.01	0.051	5.3	0.0	-0.22	0.206
1.4	172	0.00	0.051	4.7	0.0	-0.28	0.254
1.5	109	0.01	0.055	5.5	0.0	-0.33	0.268
1.6	75	0.01	0.039	1.3	0.0	-0.35	0.210
1.7	63	0.02	0.037	1.6	0.0	-0.39	0.180
1.8	42	0.03	0.037	4.8	0.0	-0.41	0.260
2.0	27	0.07	0.043	3.7	0.0	-0.41	0.240

- * Error = Owen's estimate - True posterior mode.
Adjusted Error = Error - Mean error at that ability level.
Order Effect = Owen estimate with item order reversed
- Original Owen estimate.

Table 2

RESULTS FOR SUBTEST AR*

Owen's Estimate	N	<u>Errors</u>		<u>Adjusted Errors</u>		<u>Order Effect</u>	
		Mean	S.D.	%Above 0.1	%Below -0.1	Mean	S.D.
-1.0	39	-0.06	0.084	10.3	0.0	-0.13	0.183
-0.9	48	-0.06	0.080	10.4	0.0	-0.12	0.118
-0.8	58	-0.08	0.067	3.4	0.0	-0.06	0.104
-0.7	82	-0.07	0.079	4.9	0.0	-0.08	0.095
-0.6	109	-0.07	0.051	6.4	0.0	-0.08	0.116
-0.5	126	-0.06	0.067	5.6	0.0	-0.12	0.114
-0.4	168	-0.06	0.070	3.0	0.0	-0.07	0.093
-0.3	205	-0.07	0.076	2.4	0.0	-0.09	0.095
-0.2	304	-0.07	0.059	3.6	0.0	-0.09	0.096
-0.1	316	-0.07	0.054	2.8	0.0	-0.08	0.102
0.0	310	-0.07	0.061	3.9	0.0	-0.08	0.096
0.1	295	-0.07	0.052	4.1	0.0	-0.09	0.100
0.2	297	-0.06	0.054	4.0	0.0	-0.09	0.126
0.3	344	-0.06	0.049	4.7	0.0	-0.11	0.122
0.4	361	-0.05	0.057	6.9	0.0	-0.13	0.126
0.5	342	-0.04	0.047	5.0	0.0	-0.12	0.115
0.6	362	-0.03	0.050	5.0	0.0	-0.12	0.113
0.7	391	-0.03	0.045	2.6	0.0	-0.12	0.111
0.8	365	-0.01	0.047	1.1	0.0	-0.15	0.112
0.9	282	0.00	0.049	3.5	0.0	-0.14	0.124
1.0	259	0.00	0.050	5.4	0.0	-0.17	0.149
1.1	230	0.02	0.058	8.3	0.0	-0.18	0.169
1.2	180	0.03	0.061	10.6	0.6	-0.19	0.158
1.3	149	0.03	0.057	9.4	0.0	-0.24	0.184
1.4	143	0.05	0.068	11.2	0.7	-0.24	0.196
1.5	70	0.06	0.061	12.9	0.0	-0.28	0.171
1.6	54	0.09	0.067	13.0	0.0	-0.16	0.287
1.7	54	0.09	0.083	7.4	0.0	-0.14	0.292
1.8	34	0.09	0.034	2.9	0.0	-0.15	0.266
1.9	33	0.11	0.037	3.0	0.0	-0.04	0.305
2.0	26	0.11	0.013	0.0	0.0	-0.18	0.347
2.3	28	0.13	0.024	0.0	0.0	0.28	0.074

- * Error = Owen's estimate - True posterior mode.
Adjusted Error = Error - Mean error at that ability level.
Order Effect = Owen estimate with item order reversed
- Original Owen estimate.

Table 3

RESULTS FOR SUBTEST WK*

Owen's Estimate	N	<u>Errors</u>		<u>Adjusted Errors</u>		<u>Order Effect</u>	
		Mean	S.D.	%Above 0.1	%Below -0.1	Mean	S.D.
-0.9	30	-0.01	0.094	13.3	0.0	-0.21	0.150
-0.8	39	-0.04	0.081	5.1	0.0	-0.17	0.185
-0.7	72	-0.06	0.038	2.8	0.0	-0.19	0.176
-0.6	117	-0.07	0.054	3.4	0.0	-0.21	0.185
-0.5	143	-0.08	0.050	4.2	0.0	-0.18	0.177
-0.4	186	-0.08	0.060	3.2	0.0	-0.13	0.126
-0.3	218	-0.07	0.063	1.8	0.0	-0.15	0.136
-0.2	303	-0.07	0.046	2.0	0.0	-0.12	0.119
-0.1	293	-0.06	0.054	2.4	0.0	-0.14	0.125
0.0	357	-0.05	0.046	2.5	0.0	-0.17	0.135
0.1	427	-0.04	0.064	4.7	0.0	-0.17	0.143
0.2	485	-0.04	0.051	3.9	0.2	-0.18	0.132
0.3	501	-0.04	0.060	5.6	0.2	-0.19	0.132
0.4	512	-0.04	0.056	8.2	0.0	-0.20	0.154
0.5	521	-0.03	0.061	6.9	0.2	-0.22	0.155
0.6	477	-0.03	0.057	6.7	0.0	-0.23	0.174
0.7	511	-0.03	0.055	5.3	0.0	-0.25	0.174
0.8	415	-0.02	0.054	6.3	0.0	-0.27	0.181
0.9	388	-0.01	0.058	6.2	0.0	-0.28	0.171
1.0	308	0.00	0.051	6.2	0.0	-0.28	0.170
1.1	259	0.00	0.051	4.2	0.0	-0.30	0.166
1.2	246	0.01	0.058	5.7	0.0	-0.36	0.197
1.3	189	0.01	0.052	3.2	0.0	-0.36	0.192
1.4	108	0.02	0.047	4.6	0.9	-0.38	0.221
1.5	98	0.02	0.048	7.1	0.0	-0.38	0.208
1.6	62	0.06	0.066	4.8	0.0	-0.35	0.311
1.7	57	0.05	0.043	3.5	0.0	-0.21	0.243
1.8	34	0.07	0.049	5.9	0.0	-0.31	0.406
1.9	39	0.09	0.023	0.0	0.0	-0.05	0.157

- * Error = Owen's estimate - True posterior mode.
Adjusted Error = Error - Mean error at that ability level.
Order Effect = Owen estimate with item order reversed
- Original Owen estimate.

Table 4

RESULTS FOR SUBTEST PC*

Owen's Estimate	N	<u>Errors</u>		<u>Adjusted Errors</u>		<u>Order Effect</u>	
		Mean	S.D.	%Above 0.1	%Below -0.1	Mean	S.D.
-1.1	36	-0.02	0.145	16.7	11.1	-0.19	0.142
-1.0	39	-0.06	0.096	15.4	0.0	-0.13	0.137
-0.9	55	-0.03	0.172	14.5	9.1	-0.12	0.101
-0.8	44	-0.06	0.145	13.6	0.0	-0.10	0.112
-0.7	79	-0.05	0.163	17.7	6.3	-0.11	0.112
-0.6	89	-0.09	0.106	13.5	1.1	-0.11	0.120
-0.5	126	-0.10	0.083	5.6	0.0	-0.10	0.098
-0.4	170	-0.11	0.100	4.7	0.0	-0.12	0.101
-0.3	239	-0.13	0.062	4.2	0.0	-0.12	0.124
-0.2	259	-0.13	0.077	7.3	1.2	-0.13	0.128
-0.1	282	-0.12	0.076	9.9	2.5	-0.12	0.137
0.0	455	-0.12	0.058	2.0	3.1	-0.10	0.120
0.1	462	-0.12	0.054	1.9	2.2	-0.11	0.116
0.2	414	-0.12	0.051	3.1	0.7	-0.12	0.120
0.3	450	-0.11	0.058	5.8	0.4	-0.13	0.126
0.4	458	-0.09	0.073	14.0	0.7	-0.14	0.132
0.5	371	-0.08	0.069	11.6	0.0	-0.16	0.131
0.6	394	-0.05	0.075	17.0	0.0	-0.18	0.134
0.7	348	-0.05	0.066	12.6	0.0	-0.17	0.123
0.8	319	-0.03	0.072	11.0	2.8	-0.21	0.106
0.9	270	0.00	0.072	10.0	4.4	-0.21	0.141
1.0	213	0.03	0.073	10.3	5.2	-0.24	0.157
1.1	217	0.03	0.069	7.8	0.9	-0.28	0.155
1.2	129	0.05	0.087	14.0	0.0	-0.31	0.210
1.3	141	0.08	0.094	14.9	0.0	-0.34	0.265
1.4	118	0.15	0.109	25.4	15.3	-0.35	0.304
1.5	114	0.12	0.080	10.5	0.0	-0.39	0.268
1.7	86	0.14	0.014	0.0	0.0	-0.58	0.250
1.8	89	0.17	0.016	0.0	0.0	-0.64	0.231
1.9	40	0.19	0.022	0.0	0.0	-0.85	0.345
2.1	28	0.16	0.032	3.6	0.0	0.29	0.056
2.2	28	0.21	0.028	0.0	0.0	0.17	0.022

- * Error = Owen's estimate - True posterior mode.
Adjusted Error = Error - Mean error at that ability level.
Order Effect = Owen estimate with item order reversed
- Original Owen estimate.

Table 5

RESULTS FOR SUBTEST AI*

Owen's Estimate	N	<u>Errors</u>		<u>Adjusted Errors</u>		<u>Order Effect</u>	
		Mean	S.D.	%Above 0.1	%Below -0.1	Mean	S.D.
-1.2	51	-0.03	0.103	13.7	13.7	-0.21	0.151
-1.1	93	-0.08	0.067	7.5	1.1	-0.22	0.191
-1.0	116	-0.08	0.098	7.8	5.2	-0.22	0.180
-0.9	181	-0.11	0.064	5.0	2.2	-0.18	0.159
-0.8	187	-0.11	0.069	4.3	1.6	-0.14	0.153
-0.7	229	-0.11	0.063	5.7	0.4	-0.14	0.167
-0.6	259	-0.10	0.061	4.2	0.0	-0.14	0.157
-0.5	286	-0.10	0.053	6.3	0.0	-0.13	0.162
-0.4	296	-0.08	0.078	5.1	1.4	-0.14	0.137
-0.3	309	-0.09	0.051	2.6	1.6	-0.14	0.150
-0.2	274	-0.09	0.063	6.9	1.8	-0.14	0.145
-0.1	232	-0.10	0.072	9.5	3.4	-0.13	0.150
0.0	235	-0.10	0.060	8.1	2.1	-0.14	0.147
0.1	233	-0.10	0.065	6.4	0.9	-0.13	0.148
0.2	254	-0.08	0.063	7.1	2.4	-0.14	0.151
0.3	243	-0.08	0.066	7.4	3.3	-0.16	0.165
0.4	273	-0.07	0.069	8.8	5.9	-0.18	0.176
0.5	283	-0.07	0.070	6.7	7.8	-0.19	0.172
0.6	296	-0.07	0.086	10.8	11.1	-0.20	0.172
0.7	281	-0.06	0.091	13.2	12.8	-0.22	0.187
0.8	297	-0.06	0.095	11.4	13.5	-0.23	0.201
0.9	196	-0.06	0.090	12.2	15.3	-0.29	0.260
1.0	208	-0.04	0.080	13.5	10.6	-0.33	0.254
1.1	219	-0.02	0.091	12.8	12.8	-0.37	0.289
1.2	178	-0.01	0.087	10.7	11.8	-0.39	0.256
1.3	146	0.00	0.101	15.1	11.6	-0.47	0.299
1.4	118	0.01	0.097	18.6	13.6	-0.51	0.328
1.5	88	0.05	0.091	13.6	12.5	-0.54	0.376
1.6	71	0.04	0.112	19.7	16.9	-0.57	0.303
1.7	58	0.08	0.121	13.8	20.7	-0.52	0.309
1.8	43	0.08	0.113	18.6	25.6	-0.47	0.329
1.9	27	0.12	0.100	11.1	18.5	-0.51	0.373

- * Error = Owen's estimate - True posterior mode.
Adjusted Error = Error - Mean error at that ability level.
Order Effect = Owen estimate with item order reversed
- Original Owen estimate.

Table 6

RESULTS FOR SUBTEST SI*

Owen's Estimate	N	<u>Errors</u>		<u>Adjusted Errors</u>		<u>Order Effect</u>	
		Mean	S.D.	%Above	%Below	Mean	S.D.
		-----	-----	0.1	-0.1	-----	-----
-1.6	64	0.03	0.125	14.1	0.0	-0.33	0.273
-1.5	62	0.03	0.175	11.3	0.0	-0.33	0.297
-1.4	98	-0.01	0.081	10.2	0.0	-0.34	0.253
-1.3	128	-0.02	0.087	10.2	0.0	-0.32	0.295
-1.2	143	-0.03	0.086	9.1	0.0	-0.33	0.285
-1.1	144	-0.03	0.113	7.6	0.7	-0.32	0.314
-1.0	170	-0.04	0.100	8.2	1.2	-0.31	0.321
-0.9	176	-0.07	0.073	5.7	0.0	-0.26	0.295
-0.8	147	-0.07	0.085	6.1	0.0	-0.30	0.292
-0.7	197	-0.08	0.073	4.6	0.5	-0.22	0.315
-0.6	184	-0.08	0.052	2.7	0.0	-0.14	0.280
-0.5	205	-0.08	0.096	0.5	0.0	-0.14	0.237
-0.4	232	-0.08	0.038	1.3	0.0	-0.12	0.248
-0.3	203	-0.08	0.039	2.5	0.0	-0.11	0.224
-0.2	246	-0.07	0.047	3.3	0.4	-0.09	0.212
-0.1	241	-0.07	0.041	2.5	0.0	-0.09	0.206
0.0	256	-0.06	0.039	2.7	0.8	-0.12	0.208
0.1	253	-0.05	0.048	2.0	1.2	-0.14	0.197
0.2	329	-0.04	0.054	3.6	4.3	-0.17	0.208
0.3	326	-0.03	0.061	4.9	5.8	-0.15	0.205
0.4	355	-0.03	0.069	3.9	8.7	-0.22	0.228
0.5	349	-0.03	0.085	7.4	13.2	-0.25	0.258
0.6	391	-0.02	0.086	7.7	13.8	-0.26	0.268
0.7	356	-0.04	0.094	13.5	17.4	-0.30	0.306
0.8	318	-0.04	0.095	12.9	14.5	-0.29	0.321
0.9	247	-0.05	0.085	10.9	8.5	-0.28	0.324
1.0	182	-0.05	0.080	13.7	8.2	-0.31	0.318
1.1	132	-0.05	0.070	9.8	6.8	-0.29	0.334
1.2	82	-0.04	0.068	6.1	6.1	-0.29	0.344
1.3	37	-0.06	0.071	8.1	5.4	-0.26	0.337
1.4	49	-0.05	0.079	10.2	4.1	-0.29	0.375
1.5	46	-0.03	0.076	8.7	6.5	-0.47	0.315
1.6	45	0.00	0.091	15.6	13.3	-0.39	0.391
1.7	39	0.01	0.107	12.8	12.8	-0.36	0.441
1.8	37	0.05	0.134	10.8	21.6	-0.37	0.478

- * Error = Owen's estimate - True posterior mode.
Adjusted Error = Error - Mean error at that ability level.
Order Effect = Owen estimate with item order reversed
- Original Owen estimate.

Table 7

RESULTS FOR SUBTEST MK*

Owen's Estimate	N	<u>Errors</u>		<u>Adjusted Errors</u>		<u>Order Effect</u>	
		Mean	S.D.	%Above	%Below	Mean	S.D.
		-----	-----	0.1	-0.1	-----	-----
-0.9	29	-0.07	0.041	3.4	0.0	-0.15	0.191
-0.8	42	-0.06	0.113	4.8	0.0	-0.16	0.161
-0.7	96	-0.07	0.048	7.3	0.0	-0.12	0.174
-0.6	122	-0.09	0.045	1.6	0.0	-0.11	0.150
-0.5	171	-0.09	0.042	1.2	0.0	-0.06	0.132
-0.4	170	-0.09	0.048	2.4	0.0	-0.04	0.136
-0.3	180	-0.09	0.045	2.2	0.0	-0.02	0.161
-0.2	185	-0.08	0.039	1.1	0.0	0.00	0.156
-0.1	187	-0.06	0.051	2.1	0.0	0.02	0.179
0.0	229	-0.06	0.049	3.5	0.0	-0.04	0.150
0.1	263	-0.05	0.052	4.2	0.0	-0.05	0.142
0.2	257	-0.03	0.063	5.4	0.0	-0.10	0.117
0.3	294	-0.02	0.058	4.4	0.0	-0.13	0.114
0.4	354	-0.03	0.053	4.5	0.0	-0.15	0.117
0.5	363	-0.02	0.061	7.4	0.3	-0.18	0.122
0.6	467	-0.02	0.068	8.6	1.1	-0.16	0.111
0.7	472	-0.03	0.069	6.1	0.4	-0.19	0.116
0.8	441	-0.03	0.068	6.6	0.2	-0.20	0.130
0.9	419	-0.02	0.060	5.7	0.0	-0.20	0.139
1.0	354	-0.02	0.053	4.5	0.0	-0.21	0.142
1.1	341	0.00	0.056	5.9	0.0	-0.26	0.173
1.2	299	0.01	0.061	5.0	0.0	-0.29	0.176
1.3	300	0.00	0.045	3.0	0.0	-0.30	0.183
1.4	190	0.01	0.051	3.2	0.0	-0.32	0.217
1.5	167	0.02	0.050	6.6	0.0	-0.32	0.207
1.6	111	0.01	0.043	1.8	0.0	-0.32	0.186
1.7	64	0.04	0.062	7.8	0.0	-0.19	0.311
1.8	37	0.04	0.053	5.4	0.0	-0.10	0.378
1.9	36	0.06	0.056	11.1	0.0	-0.18	0.343

- * Error = Owen's estimate - True posterior mode.
Adjusted Error = Error - Mean error at that ability level.
Order Effect = Owen estimate with item order reversed
- Original Owen estimate.

Table 8

RESULTS FOR SUBTEST MC*

Owen's Estimate	N	<u>Errors</u>		<u>Adjusted Errors</u>		<u>Order Effect</u>	
		Mean	S.D.	%Above	%Below	Mean	S.D.
		-----	-----	0.1	-0.1	-----	-----
-1.5	34	0.02	0.087	5.9	0.0	-0.16	0.175
-1.4	46	-0.02	0.117	4.3	0.0	-0.09	0.198
-1.3	52	-0.03	0.137	9.6	0.0	-0.07	0.186
-1.2	68	-0.03	0.157	11.8	0.0	-0.10	0.215
-1.1	81	-0.05	0.158	13.6	0.0	-0.06	0.206
-1.0	85	-0.03	0.174	22.4	16.5	-0.11	0.191
-0.9	108	-0.06	0.160	15.7	8.3	-0.06	0.172
-0.8	113	-0.07	0.155	20.4	9.7	-0.04	0.160
-0.7	144	-0.08	0.167	13.9	10.4	-0.05	0.168
-0.6	146	-0.10	0.137	9.6	4.8	-0.05	0.162
-0.5	199	-0.11	0.158	7.5	4.0	-0.06	0.153
-0.4	218	-0.11	0.113	8.7	3.7	-0.03	0.147
-0.3	252	-0.12	0.105	5.2	3.2	-0.04	0.179
-0.2	285	-0.11	0.110	5.3	1.1	-0.05	0.159
-0.1	332	-0.11	0.081	6.0	0.0	0.04	0.140
0.0	364	-0.09	0.087	9.6	1.1	-0.08	0.154
0.1	450	-0.08	0.085	8.0	0.4	-0.07	0.146
0.2	447	-0.07	0.080	5.6	0.2	-0.06	0.161
0.3	555	-0.06	0.061	7.2	0.0	-0.10	0.170
0.4	455	-0.06	0.055	5.3	0.0	-0.16	0.158
0.5	381	-0.05	0.062	7.1	0.3	-0.17	0.169
0.6	382	-0.03	0.074	8.4	1.6	-0.14	0.169
0.7	386	-0.02	0.072	8.5	1.3	-0.20	0.173
0.8	298	-0.01	0.076	10.4	2.3	-0.23	0.185
0.9	214	0.02	0.084	11.2	1.4	-0.28	0.192
1.0	186	0.03	0.087	14.5	0.5	-0.29	0.202
1.1	120	0.04	0.097	20.8	2.5	-0.30	0.209
1.2	102	0.02	0.093	12.7	0.0	-0.20	0.213
1.3	66	0.04	0.113	15.2	4.5	-0.31	0.198
1.4	48	0.05	0.100	18.8	6.3	-0.29	0.277
1.6	26	0.07	0.074	3.8	0.0	-0.18	0.344

- * Error = Owen's estimate - True posterior mode.
Adjusted Error = Error - Mean error at that ability level.
Order Effect = Owen estimate with item order reversed
- Original Owen estimate.

Table 9

RESULTS FOR SUBTEST EI*

Owen's Estimate	N	<u>Errors</u>		<u>Adjusted Errors</u>		<u>Order Effect</u>	
		Mean	S.D.	%Above 0.1	%Below -0.1	Mean	S.D.
-1.5	31	-0.01	0.071	12.9	0.0	-0.18	0.174
-1.4	41	-0.01	0.097	14.6	0.0	-0.18	0.153
-1.3	57	-0.03	0.055	8.8	0.0	-0.25	0.174
-1.2	108	-0.03	0.088	3.7	0.0	-0.22	0.178
-1.1	117	-0.04	0.050	6.0	0.0	-0.21	0.193
-1.0	171	-0.04	0.075	5.3	0.0	-0.21	0.205
-0.9	198	-0.05	0.055	2.5	0.0	-0.22	0.180
-0.8	261	-0.05	0.092	3.8	0.0	-0.20	0.162
-0.7	252	-0.06	0.043	3.2	0.0	-0.15	0.169
-0.6	347	-0.06	0.065	3.5	0.0	-0.14	0.172
-0.5	245	-0.07	0.065	4.5	0.0	-0.12	0.160
-0.4	249	-0.06	0.059	2.4	0.0	-0.07	0.143
-0.3	285	-0.06	0.057	2.5	0.0	-0.07	0.146
-0.2	321	-0.07	0.051	2.8	0.0	-0.08	0.145
-0.1	303	-0.06	0.048	3.3	0.0	-0.09	0.150
0.0	343	-0.05	0.055	3.8	0.0	-0.10	0.150
0.1	252	-0.05	0.054	4.8	0.0	-0.12	0.153
0.2	266	-0.03	0.072	6.0	0.0	-0.14	0.145
0.3	214	-0.04	0.071	7.0	0.0	-0.14	0.164
0.4	243	-0.02	0.071	8.6	1.6	-0.13	0.139
0.5	187	-0.03	0.072	13.9	1.6	-0.13	0.148
0.6	204	-0.01	0.075	11.3	5.9	-0.16	0.154
0.7	174	-0.03	0.063	6.9	2.9	-0.16	0.185
0.8	243	-0.04	0.055	4.9	0.4	-0.21	0.175
0.9	151	-0.03	0.079	9.9	2.0	-0.25	0.211
1.0	160	-0.05	0.080	6.9	1.3	-0.28	0.203
1.1	144	-0.06	0.061	6.9	0.0	-0.31	0.196
1.2	131	-0.04	0.078	9.9	0.0	-0.28	0.154
1.3	60	-0.04	0.074	6.7	0.0	-0.29	0.142
1.4	62	-0.03	0.046	4.8	0.0	-0.30	0.154
1.5	45	-0.02	0.047	4.4	0.0	-0.30	0.206
1.6	43	0.01	0.054	7.0	0.0	-0.42	0.203
1.7	35	0.03	0.055	5.7	0.0	-0.39	0.245
1.8	26	0.00	0.045	3.8	0.0	-0.32	0.118
1.9	32	0.04	0.056	9.4	0.0	-0.35	0.248

- * Error = Owen's estimate - True posterior mode.
Adjusted Error = Error - Mean error at that ability level.
Order Effect = Owen estimate with item order reversed
- Original Owen estimate.

REFERENCES

- [1] Owen, Roger J. "A Bayesian Sequential Procedure for Quantal Response in the Context of Adaptive Mental Testing." Journal of the American Statistical Association(June 1975): 351-356
- [2] Rehab Group, Inc., "Predictive Utility Evaluation of Computerized Adaptive Testing: Results of the Navy Research," by Susan B. Hardwicke and Kenneth D. White, Dec 1983
- [3] Discussion with J. Bradford Sympson, Navy Personnel Research and Development Center, Nov 1985

END

2-87.

DTIC